Broj rada: R5-03
DOI:10.46793/CIRED2026.R5-03_197

PROGNOZA OPTEREĆENJA NISKONAPONSKOG IZVODA UPOTREBOM MAŠINSKOG UČENJA

LOAD FORECASTING OF MEDIUM-VOLTAGE FEEDER LOAD USING MACHINE LEARNING

Aleksa ALAGIĆ, Elektromreže Srbije/ Elektrotehnički fakultetu u Beogradu, Srbija
Darko ŠOŠIĆ, Elektrotehnički fakultet u Beogradu, Srbija

KRATAK SADRŽAJ

Usled sve većeg učešća obnovljivih izvora energije, kao i kontinuiranog rasta potrošnje električne energije, balansiranje elektroenergetskog sistema postaje sve složeniji zadatak. Razvoj i masovna primena pametnih brojila omogućili su prikupljanje velike količine podataka, čime prognoziranje potrošnje postaje ne samo moguće, već i neophodno. Tačna prognoza opterećenja ima ključnu ulogu kako u svakodnevnom balansiranju sistema, tako i u planiranju mreže i izdavanju dozvola za izgradnju novih kapaciteta obnovljivih izvora energije. U ovom radu razmatra se prognoza opterećenja niskonaponske mreže primenom mašinskog učenja. Korišćeni su realni podaci merenja iz distributivne mreže, čime je omogućena verodostojnija analiza potrošnje i preciznije određivanje odstupanja između prognoziranih i stvarnih vrednosti. Skup podataka podeljen je na deo za obučavanje, validaciju i testiranje, u cilju postizanja pouzdanih i stabilnih rezultata. Sirovi merni podaci prethodno su obrađeni kako bi se uklonili šumovi i nepravilni unosi, kao i popunile praznine nastale usled nedostajućih merenja, čime su formirani kvalitetni ulazni podaci za model. Dobijeni rezultati ukazuju na visok potencijal primene mašinskog učenja u oblasti prognoze potrošnje električne energije i predstavljaju dobru osnovu za dalja istraživanja u savremenoj elektroenergetici. Imajući u vidu brz napredak tehnologije, može se zaključiti da će se u budućnosti veštačke neuralne mreže sve intenzivnije koristiti u planiranju i upravljanju elektroenergetskim sistemima. Dalji pravac istraživanja može obuhvatiti određivanje optimalnih lokacija za ugradnju fotonaponskih sistema, čime bi se dodatno smanjilo opterećenje niskonaponske distributivne mreže.

Ključne reči: prognoza opterećenja, mašinsko učenje, pametna brojila, veštačka inteligencija, niskonaponska distributivna mreža

ABSTRACT

Due to the increasing penetration of renewable energy sources and the continuous growth of electricity consumption, balancing the power system is becoming an increasingly complex task. The development and widespread deployment of smart meters have enabled the collection of large volumes of data, making load forecasting not only feasible but also essential. Accurate load forecasting plays a key role not only in day-to-day system balancing, but also in network planning and in the process of granting permits for the construction of new renewable energy capacities. This paper addresses load forecasting in low-voltage distribution networks using machine learning. Real measurement data obtained from the distribution network are used,

* Aleksa Alagić alagic.aleksa7@gmail.com; Darko Šošić sosic@etf.rs

enabling a more reliable analysis of electricity consumption and a more accurate assessment of deviations between forecasted and actual values. The dataset is divided into training, validation, and testing subsets in order to ensure reliable and stable model performance. Prior to model training, the raw measurement data are preprocessed to remove noise and invalid entries, as well as to fill gaps caused by missing measurements, thereby forming high-quality input data for the model. The obtained results indicate a high potential for the application of machine learning in electricity load forecasting and provide a solid foundation for further research in modern power systems. Considering the rapid technological development, it can be concluded that artificial neural networks will be increasingly used in the planning and operation of power systems in the future. Future work may include the determination of optimal locations for photovoltaic system installation, which could further reduce the loading of low-voltage distribution networks.

Keywords: load forecasting, machine learning, smart meters, artificial intelligence, low-voltage distribution network

UVOD

Usled razvoja mašinskog učenja, postalo je ne samo poželjno već i neophodno unaprediti tradicionalne metode i zameniti ih savremenim pristupima. Kvalitetna prognoza potrošnje od ključnog je značaja za optimalno angažovanje proizvodnih jedinica, čime se direktno utiče na smanjenje troškova balansiranja sistema. Istovremeno, razvoj savremenih metoda za procenu stanja distributivnih mreža dobija na značaju usled sve većeg učešća distribuiranih izvora energije. Savremene distributivne mreže, naročito na nižim naponskim nivoima, često ne raspolažu dovoljnim brojem uređaja za daljinsko merenje i nadzor, što otežava pouzdano praćenje njihovog rada. Zbog toga je neophodno primeniti inovativne pristupe koji omogućavaju efikasno upravljanje sistemom uprkos ograničenoj dostupnosti i kvalitetu podataka. Ovakvi uslovi nameću potrebu za razvojem naprednih metoda procene stanja koje se oslanjaju na oskudna i nepouzdana merenja.

Povećano učešće obnovljivih izvora energije na distributivnom nivou, kao i promene u potrošačkim navikama u savremenom društvu, doveli su do promene karakteristika dijagrama opterećenja, uključujući pomeranje vršnih opterećenja i pojavu dodatnih pikova, na primer u dnevnim časovima oko podneva. U skladu sa tim, ovaj rad se bavi predviđanjem potrošnje primenom metode *gradient boosting*. Ova metoda ima široku primenu, od predviđanja potrošnje [1], preko medicine [2] pa do finansija [3].

Pored ove metode, postoji i niz drugih pristupa zasnovanih na mašinskom učenju koji su našli primenu u energetici, kao što su prognoza proizvodnje iz obnovljivih izvora energije [4], predikcija kvarova [5], kao i procena stanja sistema [6].

1 PROGNOZA POTROŠNJE U PROŠLOSTI

Pre nego što se pristupi detaljnoj analizi metode *gradient boosting*, potrebno je dati kratak osvrt na matematičke modele koji su se tradicionalno koristili za prognozu potrošnje električne energije. Iako su ovi modeli decenijama predstavljali osnovu analize vremenskih serija, njihov značaj se danas postepeno smanjuje sa razvojem metoda zasnovanih na mašinskom učenju i veštačkoj inteligenciji.

Razvoj metoda za analizu vremenskih serija doveo je do široke primene klasičnih statističkih modela, među kojima se posebno izdvajaju ARIMA i SARIMA pristupi. Njihova popularnost proističe iz jasno definisane strukture i relativno jednostavne interpretacije rezultata. Međutim, u realnim sistemima, kao što je elektroenergetska potrošnja, često se javljaju izražene nelinearnosti i uticaji spoljašnjih faktora (npr. meteorološki uslovi ili promene u obrascima ponašanja korisnika), što smanjuje efikasnost ovih modela. Sa druge strane, savremene tehnike mašinskog i dubokog učenja omogućavaju rad sa velikim količinama heterogenih podataka i bolje hvatanje složenih zavisnosti, čime se unapređuje tačnost prognoza. Ipak, tradicionalni statistički modeli i dalje imaju značajnu ulogu, kako zbog teorijske utemeljenosti, tako i kao referentna osnova za evaluaciju naprednijih pristupa [7].

Jedan od ključnih predstavnika ove klase modela je ARIMA model, razvijen u okviru Box–Jenkins metodologije, koji je namenjen analizi i prognozi vremenskih serija sa izraženom međuzavisnošću između uzastopnih vrednosti. Njegova primena obuhvata različite oblasti, uključujući modelovanje klimatskih i hidroloških pojava, poput temperature, padavina i protoka.

ARIMA model se standardno označava kao ARIMA(p, d, q), gde parametar p definiše red autoregresivne komponente, d označava broj potrebnih diferenciranja za postizanje stacionarnosti, dok q predstavlja red komponente pokretnog proseka. Autoregresivni deo modela uspostavlja vezu između trenutne vrednosti i prethodnih vrednosti vremenske serije, što omogućava identifikaciju obrazaca na osnovu istorijskih podataka. Komponenta pokretnog proseka uzima u obzir uticaj prethodnih odstupanja (grešaka) na trenutnu vrednost, dok integrisana komponenta obezbeđuje transformaciju nestacionarne serije u stacionarnu, što je osnovni preduslov za primenu modela [8].

$$Y' = c + \phi_1 \cdot y'_{t-1} + \cdots + \phi_p \cdot y'_{t-p} + \theta_1 \cdot \varepsilon'_{t-1} + \cdots + \theta_q \cdot \varepsilon'_{t-q} + \varepsilon_t \quad (1)$$

Autoregresivna komponenta (AR) opisana je izrazom

$$\phi_1 \cdot y'_{t-1} + \cdots + \phi_p \cdot y'_{t-p} \quad (2)$$

Komponenta pokretnog proseka (MA) opisana je izrazom gde ε_t predstavlja slučajnu grešku (engl. error term)

$$\theta_1 \cdot \varepsilon'_{t-1} + \cdots + \theta_q \cdot \varepsilon'_{t-q} \quad (3)$$

Integrisana komponenta (I), definisana parametrom d, određuje stepen diferenciranja koji je potrebno primeniti kako bi vremenska serija postala stacionarna, što je osnovni preduslov za primenu ARIMA modela.

Model se teško snalazi sa obuhvatanjem sezonskih obrazaca ponašanja, osim ukoliko se ne proširi u formu sezonskog ARIMA (SARIMA) modela. Da bi se unapredila tačnost modelovanja vremenskih serija koje pokazuju sezonske obrasce, razvijen je SARIMA model kao proširenje osnovnog ARIMA pristupa. Njegova osnovna svrha je da omogući bolju reprezentaciju podataka kod kojih se uočavaju ponavljajući obrasci u određenim vremenskim intervalima. Za razliku od standardnog ARIMA modela, SARIMA uvodi dodatne parametre koji opisuju sezonsko ponašanje podataka. Na taj način moguće je posebno modelovati sezonske efekte, pored uobičajenih nesezonskih komponenti. U okviru ovog pristupa, koriste se sezonski oblici autoregresije, diferenciranja i pokretnog proseka, što omogućava modelu da

uoči pravilnosti koje se ponavljaju kroz određene periode. Ovakav način modelovanja posebno je koristan u slučajevima kada podaci pokazuju jasne ciklične obrasce, kao što su mesečne promene u prodaji, kvartalne promene ekonomskih pokazatelja ili godišnje varijacije klimatskih veličina. U takvim situacijama, jednostavni modeli često nisu dovoljni jer ne mogu adekvatno da opišu sezonsku strukturu podataka. SARIMA model se u literaturi najčešće zapisuje u obliku SARIMA(p, d, q)(P, D, Q, s), gde prvi deo parametara opisuje nesezonske komponente, dok drugi deo predstavlja sezonske komponente modela. Parametar s označava dužinu jednog sezonskog ciklusa, odnosno broj vremenskih koraka nakon kojih se obrazac ponavlja. Ovakva struktura omogućava modelu da istovremeno uči i kratkoročne zavisnosti i dugoročne sezonske obrasce, što ga čini znatno fleksibilnijim u odnosu na osnovni ARIMA model [9].

2 PROGNOZA POTROŠNJE KORIŠĆENJEM MAŠINSKOG UČENJE

Savremeni algoritmi mašinskog učenja često se zasnivaju na ideji kombinovanja većeg broja jednostavnijih modela kako bi se dobio jedan snažan prediktor. U tom kontekstu razvijene su metode poput *bagging*-a i *boosting*-a. Dok *bagging* formira konačnu predikciju agregacijom rezultata modela treniranih na različitim *bootstrap* uzorcima, *boosting* pristup gradi modele sekvencijalno, pri čemu se u svakoj iteraciji veća pažnja posvećuje primerima koji su prethodno lošije predviđeni. Jedan od najpoznatijih predstavnika boosting metoda je AdaBoost, koji se može interpretirati kao postupak minimizacije eksponencijalne funkcije gubitka. Ovaj koncept je dalje proširen kroz pristup poznat kao *gradient boosting*, gde se učenje novog modela zasniva na aproksimaciji gradijenta funkcije gubitka iz prethodne iteracije, čime se omogućava optimizacija proizvoljnih funkcija gubitka. *Gradient boosting* stabla predstavljaju ansambl pristup u kojem se modeli grade postepeno, tako što svako novo stablo pokušava da koriguje greške prethodno formiranog ansambla. Neka je dat skup podataka sa određenim brojem primera i karakteristika $D = \{(x_i, y_i) \mid i \in \{1, .., n\},\ x_i \in \mathbb{R}^m,\ y_i \in \mathbb{R}\}$. Predikcija ansambla dobija se sumiranjem izlaza svih pojedinačnih stabala. U procesu izgradnje narednog stabla optimizuje se funkcija cilja koja uključuje komponentu gubitka i regularizacioni član [10].

$$L = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{\,K} + f_{K+1}(x_i)\right) + \theta\,(f_{K+1}) \tag{4}$$

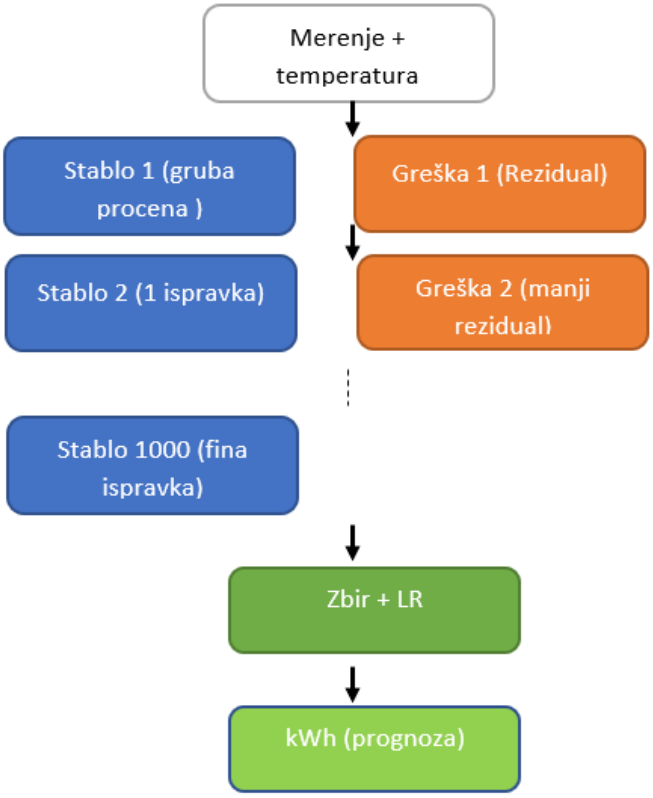
Funkcija gubitka zavisi od odstupanja između stvarnih i predviđenih vrednosti, dok regularizacija ograničava složenost modela i time doprinosi smanjenju prekomernog prilagođavanja. Ovo praktično znači:

- Ansambl suma ($\sum f_i$): Finalna prognoza potrošnje nije rezultat jednog "pametnog" stabla, već zbira hiljada malih ispravki. Prvo stablo da grubu procenu, a svako sledeće dodaje sitnu korekciju (npr. +2 MW ili -1.5 MW).

- Prvi i drugi moment: Ovo se odnosi na Tejlorov red. XGBoost koristi i prvi izvod (gradijent) i drugi izvod (Hessian) funkcije gubitka. To omogućava algoritmu da mnogo brže "shvati" u kom pravcu treba da menja model da bi smanjio grešku.

- Regularizacija (θ): Ovo je "kočnica". Bez ovoga, model bi postao toliko kompleksan da bi savršeno predvideo prošlost (trening podatke), ali bi bio neupotrebljiv za budućnost. Ona kažnjava stabla koja imaju previše listova ili prevelike težine.

Sam proces treniranja stabla zasniva se na iterativnom proširivanju strukture stabla kroz izbor optimalnih podela, pri čemu pretraga svih mogućih tačaka razdvajanja predstavlja jedan od ključnih računskih izazova, naročito kod velikih skupova podataka. *Gain function* (Funkcija dobitka): To je matematički kriterijum koji kaže: "Ako podelim podatke na ovom mestu (npr. temperatura > 25°C), koliko će moje predviđanje postati preciznije?". Algoritam bira onaj rez koji najviše smanjuje grešku.



Slika 1 – Način rada *gradient boost*-a

3 GREŠKE PROGNOZE

U cilju evaluacije performansi modela korišćene su standardne metrike greške: srednja apsolutna greška (MAE), srednja apsolutna procentualna greška (MAPE) i koeficijent determinacije (R^2). Ove metrike omogućavaju kvantitativnu procenu odstupanja između stvarnih i predviđenih vrednosti, kao i ukupne sposobnosti modela da objasni varijabilnost podataka.

Srednja apsolutna greška (MAE) predstavlja prosečnu vrednost apsolutnih razlika između stvarnih i predviđenih vrednosti. Ova metrika daje intuitivnu meru prosečne greške u istim jedinicama kao posmatrana veličina i ne pravi razliku između pozitivnih i negativnih odstupanja:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

gde je y_i stvarna vrednost, $\hat{y}_i$ predviđena vrednost, a n ukupan broj uzoraka.

Srednja apsolutna procentualna greška (MAPE) izražava grešku u procentima, čime omogućava lakšu interpretaciju relativnog odstupanja modela u odnosu na stvarne vrednosti:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{6}$$

Koeficijent determinacije (R^2) meri koliko dobro model objašnjava varijansu posmatranih podataka. Njegova vrednost se kreće od 0 do 1 (uz mogućnost i negativnih vrednosti kod loših modela), pri čemu veće vrednosti ukazuju na bolje slaganje modela sa podacima:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{7}$$

gde je $\bar{y}$ srednja vrednost stvarnih podataka.

4 UČITAVANJE PODATAKA I FILTRIRANJE

Učitavanje podataka i njihovo filtriranje uvek predstavljaju jedan od najvećih izazova, autora rada je uneo i deo koji generiše praznike kako bi greška prognoze bila manja.

4.1 Filtriranje podataka

Podaci koji se mere u realnom vremenu često sadrže greške merenja, praznine i gruba odstupanja. U ovom radu primenjen je *pipeline* metoda čišćenja podataka koja se sastoji od pet uzastopnih slojeva, pri čemu svaki naredni sloj obrađuje greške koje prethodni sloj nije pokrio.

Prvi sloj primenjuje linearnu interpolaciju za kratke praznine u vremenskoj seriji, ograničene na najviše četiri uzastopna odbirka (jedan sat). Ovaj uslov sprečava unošenje artefakata pri interpolaciji dužih praznina.

> **Inicijalizacija:** vremenska serija $\{y_t\}_{t=1}^{T}$, korak uzorkovanja $\Delta t = 15min$, prag odstupanja $\sigma_{max} = 4$, maksimalna praznina za interpolaciju G_{max}
> Kreirati regularnu vremensku osu $\{t_g\}_{g=1}^{T'}$ sa korakom Δt
> Spojiti $\{y_t\}$ na $\{t_g\}$ pomoću left outer join → nedostajući odbirci dobijaju vrednost NaN

for svaki neprekidni niz NaN vrednosti dužine L sa granicama $[t_a,\ t_b]$
if $L \leq G_{max,1}$
for $t = t_a, \dots, t_b$

$$y_t \leftarrow y_{t_a-1} + \left(y_{t_b+1} - y_{t_a-1}\right) \cdot \frac{t - t_a + 1}{L + 1}$$

end
end if
end

Izlaz: $\{y_t\}$ sa popunjenim kratkim prazninama

Drugi sloj popunjava preostale praznine medijanom skupa referentnih odbiraka koji dele isti sat u danu i isti dan u nedelji sa posmatranom tačkom. Ovaj pristup čuva nedeljnu i dnevnu periodičnost signala, za razliku od globalne interpolacije.

for svaki t gde je $y_t = \text{NaN}$
$h \leftarrow$ sat\ _u\ _danu(t), $d \leftarrow$ dan\ _u\ _nedelji(t)
$\mathcal{R} \leftarrow \{y_i : sat(i) = h \ \wedge \ dan(i) = d \ \wedge \ y_i \neq \text{NaN}\}$
if $\mathcal{R} \neq 0$
$y_t \leftarrow medijana(\mathcal{R})$
end if
end
Izlaz: $\{y_t\}$ bez NaN vrednosti

Treći sloj identifikuje fizički nemoguće negativne vrednosti — koje nastaju usled grešaka u prenosu kumulativnih očitavanja brojila — i zamenjuje ih medijanom vremenskog prozora od ±96 odbiraka (±24 sata) oko posmatrane tačke.

for svaki t gde $y_t < 0$
$\mathcal{W} \leftarrow \{y_i : |i - t| \leq 96 \ \wedge \ i \neq t \ \wedge \ y_i \neq \text{NaN}\}$
$y_t \leftarrow medijana(\mathcal{W})$
end
Izlaz: $\{y_t\}$ bez negativnih vrednosti

Četvrti sloj primenjuje σ-metod za detekciju grubih grešaka. Odbirak $\boldsymbol{y_t}$ smatra se odstupanjem ako važi:

$$|\boldsymbol{y_t} - \boldsymbol{\mu}| > \boldsymbol{4\sigma}$$

gde su μ i σ srednja vrednost i standardna devijacija celog signala. Identifikovana odstupanja zamenjuju se linearnom interpolacijom u prozoru do osam odbiraka.

$\mu \leftarrow mean(\{yt\}), \sigma \leftarrow std(\{yt\})$
for svaki t
if $|y_t - \mu| > \sigma_{max} \cdot \sigma$
$y_t \leftarrow \text{NaN}$
end
end
for svaki neprekidni niz NaN vrednosti dužine L
if $L \leq G_{max,2}$
for $t = t_a, \dots, t_b$

$$y_t \leftarrow y_{t_a-1} + \left(y_{t_b+1} - y_{t_a-1}\right) \cdot \frac{t - t_a + 1}{L + 1}$$

end
end
end
$y_t \leftarrow \max(0, y_t) \quad \forall t$
Izlaz: $\{y_t\}$ bez grubih grešaka

Peti sloj primenjuje se isključivo na reaktivnu energiju, čija raspodela ima izražen desni rep usled induktivnih opterećenja. Umesto σ-metoda, vrednosti se ograničavaju na 99,9 percentil, čime se izbegava pogrešno uklanjanje legitimnih vršnih vrednosti.

Neka je $\{q_t\}_{t=1}^{T}$ serija reaktivne energije
$q_{999} \leftarrow \text{kvantil}(\{q_t\}, 0{,}999)$
for svaki t
$q_t \leftarrow \min(q_t, q_{999})$
end
izlaz:$\{q_t\}$ sa ograničenim vršnim vrednostima

4.2 Generisanje srpskih praznika

Za svaku godinu g iz posmatranog perioda formira se skup prazničnih dana koji se sastoji od tri grupe datuma: fiksni praznici, datumi vezani za verske i državne praznike i zamenski datuma za praznike koji padaju u nedelju. Praznici se generišu radi smanjivanja greške, predloženi algoritam za generisanje srpskih praznika je prikazan pomoću pseudo koda.

Inicijalizacija: skup praznika $\mathcal{P} \leftarrow \emptyset$, period $[g_{poč}, g_{kraj}]$
for $g = [g_{poč}, \ldots, g_{kraj}]$
Formirati skup fiksnih praznika za godinu g:
$\mathcal{F}_g \leftarrow \{01.01,\ 02.01,\ 07.01,\ 15.02,\ 16.02,\ 01.05,\ 02.05,\ 11.11\}_g$
Odrediti datum Vaskrsa v_g za godinu g iz unapred fiksne tabele
if $v_g \neq 0$
$\mathcal{V}_g \leftarrow \{v_g - 2, v_g, v_g + 1\}$
else
$\mathcal{V}_g \leftarrow \emptyset$
end
Formirati skup zamenskih dana $\mathcal{Z}_g \leftarrow \emptyset$
for svaki $f \in \mathcal{F}_g$
if dan\ _u\ _nedelji(f) = nedelja
$\mathcal{Z}_g \leftarrow \mathcal{Z}_g \cup \{f + 1\}$
end
$\mathcal{P} \leftarrow \mathcal{P} \cup \mathcal{F}_g \cup \mathcal{V}_g \cup \mathcal{Z}_g$
end
$\mathcal{P} \leftarrow \text{unique}(\mathcal{P})$
Izlaz: skup prazničnih dana $\mathcal{P}$ u formatu yyyy-MM-dd

Skup $\mathcal{P}$ se koristi u kasnijoj fazi *feature engineeringa* gde se za svaki odbirak *t* proverava da li datum pripada skupu $\mathcal{P}$, što se kodira kao binarna promenljiva $neradni_t \in \{0,1\}$

5 FEATURE ENGINEERING

„Cikličko kodiranje“ koje je korišćeno predstavlja standardnu tehniku u obradi vremenskih serija, kojom se periodične promenljive transformišu u dvodimenzionalni prostor korišćenjem sinusne i kosinusne funkcije. Ova transformacija je neophodna jer direktno numeričko kodiranje ne reflektuje stvarnu bliskost vrednosti na krajevima intervala, kao što su 23 i 0 sati ili decembar i januar.

Za promenljivu sa periodom T, transformacija se definiše kao:

$$x_{sin} = sin\left(\frac{2\pi x}{T}\right) \qquad x_{cos} = cos(\frac{2\pi x}{T}) \tag{8}$$

Na ovaj način, vrednosti koje su susedne u vremenskom smislu ostaju bliske i u transformisanom prostoru, što omogućava modelima da efikasnije uče sezonske i periodične obrasce u podacima [11].

Prilikom razmatranja prognoze potrošnje električne energije uočeno je da prošlost u velikoj meri utiče na sadašnjost. U ovakvim problemima, vrednosti iz prethodnih vremenskih trenutaka imaju značajan uticaj na buduće vrednosti. Konkretno, potrošnja u trenutku T u velikoj meri zavisi od potrošnje u prethodnom trenutku T−1. Ove istorijske vrednosti nazivaju se zaostaci (engl. *lag*), pri čemu T−1 predstavlja zaostatak prvog reda (*lag* 1), T−2 zaostatak drugog reda (*lag* 2) itd. Uvođenjem vrednosti iz ranijih vremenskih koraka (označenih kao *lag* a) moguće je formirati dodatne ulazne promenljive koje omogućavaju modelu da prepozna vremenske zavisnosti u podacima [12]. Na primer, posmatrajmo pojednostavljen skup podataka o dnevnoj potrošnji električne energije:

Tabela I – Dnevna potrošnja električne energije

Dan	Potršnja u kWh
1	100
2	120
3	150
4	180

Ako uvedemo zaostatak od jednog dana (*lag* 1), podaci se transformišu na sledeći način:

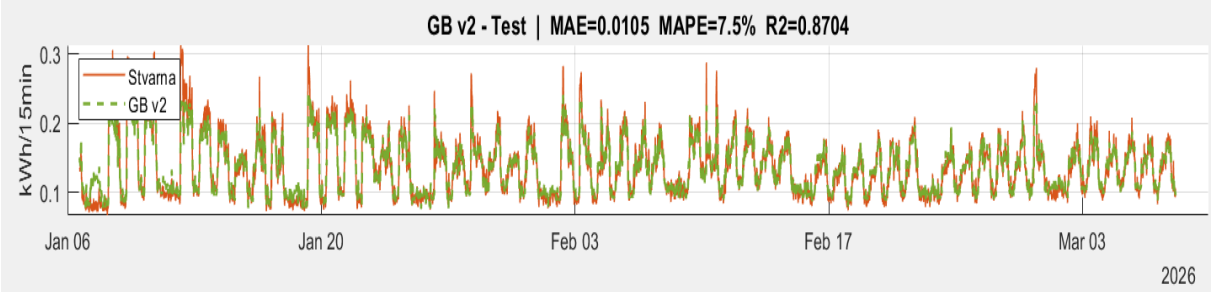
Tabela II – Podaci sa uključenim *lag* promenljivima

Dan	*Lag* 1	Potrošnja u kWh
2	100	120
3	120	150
4	150	180

Uključivanjem prethodnih vrednosti potrošnje, model dobija mogućnost da uoči obrasce u vremenu, kao što su dnevne ili sezonske varijacije. Time se poboljšava tačnost prognoze, jer se u obzir uzima zavisnost između uzastopnih merenja.

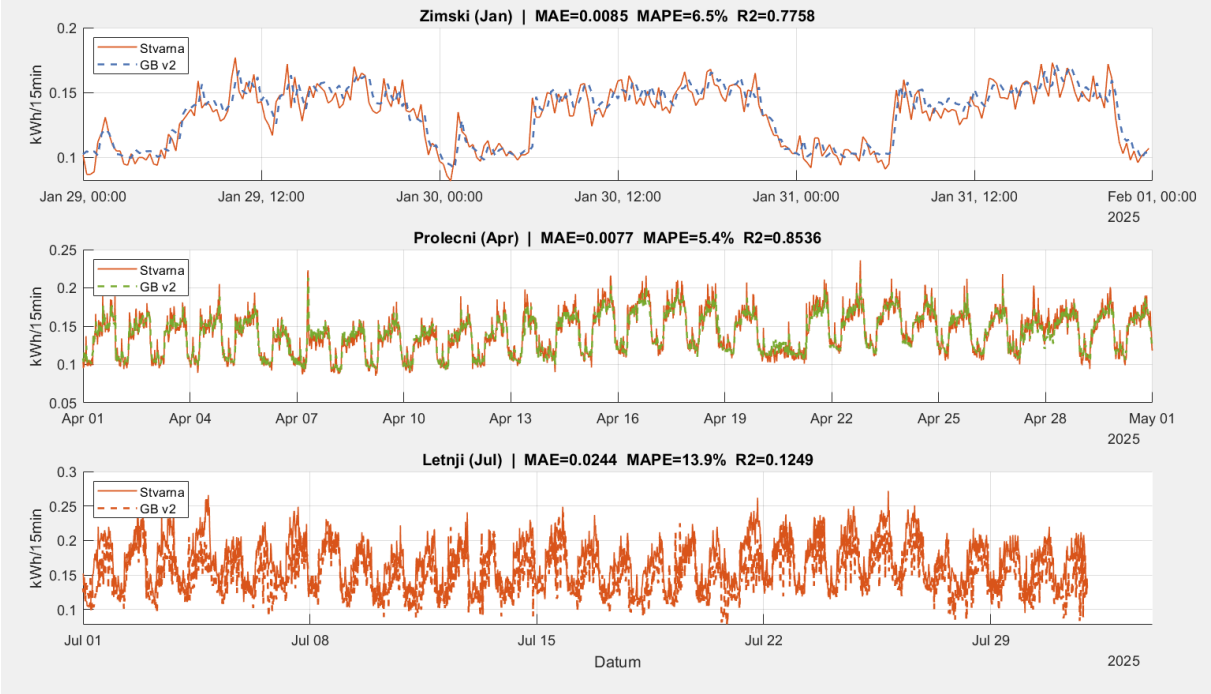
6 ULAZNI I IZLAZNI PARAMETRI

Merenja aktivne i reaktivne snage vršena su u vremenskom intervalu od 15 minuta. U modelu je primarno korišćena aktivna snaga, dok je reaktivna snaga uvedena kao dodatni ulazni parametar, s obzirom na to da je eksperimentalno uočeno smanjenje greške modela kada se ovaj podatak uključi u proces učenja. Kao referentna metrika za evaluaciju tačnosti prognoze korišćen je MAPE (Mean Absolute Percentage Error), koji predstavlja srednju apsolutnu procentualnu grešku između stvarnih i predviđenih vrednosti. Skup podataka potrošnje električne energije obuhvata period duži od godinu dana, od 01.01.2025. do 08.03.2026. godine. U početnoj fazi, za test skup su odabrani meseci januar, februar i mart 2026. godine, s obzirom na to da je na raspolaganju bila kompletna 2025. godina i dodatna tri meseca iz 2026. Međutim, ovakav pristup nije obezbeđivao reprezentativan uzorak svih sezonskih varijacija.



Slika 2 – Stvarna i predviđena potrošnja u prvom slučaju

Zbog toga je izvršena korekcija podele podataka, pri čemu su za test skup izabrani meseci januar, april i jul, koji predstavljaju zimski, prolećni i letnji period. Na ovaj način omogućena je realnija procena performansi modela u različitim sezonskim uslovima. Može se pretpostaviti da je greška u zimskom periodu manja zbog veće količine dostupnih podataka, dok je u prolećnom periodu (april) potrošnja stabilnija usled prestanka grejne sezone i manjeg broja naglih promena u ponašanju potrošača, što takođe doprinosi nižoj grešci. Nasuprot tome, u letnjem periodu dolazi do većih varijacija u potrošnji (promena temperature, korišćenje klima uređaja itd.), što rezultuje većom vrednošću greške. Pored podataka o potrošnji električne energije, u model su uključeni i standardni meteorološki parametri: datum, srednja, minimalna i maksimalna temperatura, kao i dodatni parametri – brzina vetra i atmosferski pritisak. Meteorološki podaci dostupni su na satnom nivou i vremenski su kraći u odnosu na podatke o snazi, s obzirom na to da podaci za 2026. godinu nisu bili dostupni.



Slika 3 – Stvarna i predviđena potrošnja u drugom slučaju

U cilju smanjenja vrednosti MAPE metrike, izvršena je dodatna segmentacija dana na radne i neradne, uz uključivanje državnih praznika. Na ovaj način modelu je omogućeno da bolje prepozna obrasce ponašanja potrošača u različitim režimima potrošnje. Sam postupak identifikacije i uvođenja praznika detaljno je opisan u prethodnom delu rada. Autor rada se slaže sa tezom da relativno jednostavna rešenja mogu dati veoma dobre rezultate. Ipak, uvođenjem dodatnih ulaznih parametara i unapređenjem modela postignuto je dalje poboljšanje tačnosti, pri čemu je vreme izvršavanja algoritma ostalo relativno kratko. Na kraju, posebna pažnja posvećena je analizi *lag* promenljivih, koje su prikazane u tabeli.

Tabela III – Tabela sa *lagovima*

Lag	Naziv	Vremensko pomeranje	Zašto
lag_1	15 min unazad	1 interval	Trenutni trend — šta se desilo malo pre
lag_4	1 sat unazad	4 intervala	Kratkoročni obrazac unutar istog sata
lag_96	24 sata unazad	96 intervala	Isti sat juče
lag_192	48 sati unazad	192 intervala	Isti sat prekjuče — potvrda dnevnog obrasca
lag_672	7 dana unazad	672 intervala	Isti sat prošle nedelje — nedeljni obrazac
lag_1344	14 dana unazad	1344 intervala	Dve nedelje unazad — duži nedeljni trend
lag_2688	28 dana unazad	2688 intervala	Mesec dana unazad

Takođe je razmatran njihov pojedinačni uticaj na performanse modela. Rezultati pokazuju da najveći značaj ima lag_1, što je i očekivano, s obzirom na snažnu zavisnost trenutne potrošnje

od neposredno prethodne vrednosti. Ovo pokazuje da je neke *lagove* moguće izbaciti ali da bi prikazali njihov uticaj ostavili smo ih u radu.

Tabela III – Tabela sa *lagovima* sa najvećim značajem

Najvažniji *lagovi*	Njihov uticaj
lag_1	92.2%
lag_96	0.9%
lag_4	0.7%
lag_672	0.7%

LITERATURA

[1] Hristo Ivanov Beloev, Stanislav Radikovich Saitov, Antonina Andreevna Filimonova, Natalia Dmitrievna Chichirova, Oleg Evgenievich Babikov, Iliya Krastev Iliev, „Short-Term Electrical Load Forecasting Based on XGBoost Model“
[2] Ping Xuan, Chang Sun, Tiangang Zhang, Yilin Ye, Tonghui Shen, Yihua Dong, „Gradient Boosting Decision Tree-Based Method for Predicting Interactions Between Target Genes and Drugs“
[3] Zhenya Tian, Jialiang Xiao, Haonan Feng, Yutian Wei, “Credit Risk Assessment based on Gradient Boosting Decision Tree”
[4] Natei Ermias Benti, Mesfin Diro Chaka, Addisu Gezahegn Semie, „Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects“
[5] Pathan Ilius, Mohammad Almuhaini, Muhammad Javaid, Mohammad Abido, „A Machine Learning–Based Approach for Fault Detection in Power Systems“
[6] Debottam Mukherjee, Samrat Chakraborty, Sandip Ghosh, „Power system state

Tabela IV – Uticaj *lagova*

forecasting using machine learning techniques,“
[7] Omar NgalaSarr, Moussa Fall, Mouhamadou Thiam, Mame Faty Mbaye, „Energy Consumption Forecasting Models for Smart Grids: A State-of-the-Art Review and Application Perspectives“
[8] Yuchuan Lai, David A. Dzombak, “Use of the Autoregressive Integrated Moving Average (ARIMA) Model to Forecast Near-Term Regional Temperature and Precipitation”
[9] Marcin Majka, „Seasonal Time Series Analysis: Why SARIMA Outshines ARIMA“
[10] Andreea Anghel, Nikolaos Papandreou, Thomas Parnell, Alessandro de Palma, Haralampos Pozidis, „Benchmarking and Optimization of Gradient Boosting Decision Tree Algorithms“
[11] Alice Zheng, Amanda Casari, „Feature Engineering for Machine Learning“
[12] Subrata Karmaker, „Feature Engineering in Time Series Forecasting“